A new stable basis for radial basis function interpolation

Stefano De Marchi and Gabriele Santin
Department of Mathematics
University of Padua (Italy)

Abstract

It is well-known that radial basis function interpolants suffer of bad conditioning if the basis of translates is used. In the recent work [12], the authors gave a quite general way to build stable and orthonormal bases for the native space $\mathcal{N}_\Phi(\Omega)$ associated to a kernel Φ on a domain $\Omega \subset \mathbb{R}^s$. The method is simply based on the factorization of the corresponding kernel matrix.
Starting from that setting we describe a particular basis which turns out to be orthonormal in $\mathcal{N}_\Phi(\Omega)$ and in $\ell_{2,w}(X)$, where X is a set of data sites of the domain Ω. The basis arises from a weighted singular value decomposition of the kernel matrix. This basis is also related to a discretization of the compact operator $T_\Phi : \mathcal{N}_\Phi(\Omega) \to \mathcal{N}_\Phi(\Omega)$,

$$T_\Phi[f](x) = \int_\Omega \Phi(x, y) f(y) dy \quad \forall x \in \Omega$$

and provides a connection with the continuous basis that arises from an eigen-decomposition of T_Φ. Finally, using the eigenvalues of this operator, we provide convergence estimates and stability bounds for interpolation and discrete least-squares approximation.

Keywords: Radial basis function interpolation, stability, weighted least-squares approximation.

1. Introduction

The main purpose of approximation theory is the reconstruction of a given function defined on a set $\Omega \subset \mathbb{R}^s$ from some values sampled at a finite set $X \subset \Omega$. This process is required to be convergent and stable, namely, under suitable conditions, the approximant should reproduce the original function in a chosen norm.

In this setting, the so-called *kernel methods* are of growing importance. Altough they are built to be well-posed for every data distribution, it is also well-known that the interpolation based on translates of radial basis functions or non-radial kernels is numerically unstable due to the ill-conditioning of the kernel matrix (cf. e.g. [4, 9]).
Several approaches have been studied to assure a fast convergence and stable computations. They are mainly based on different aspects of the approximation process by kernels: from the optimization of certain parameters and the search of convenient data sets, to the numerical stabilization of the underlying linear system. Other approaches are based on a change of basis for the relevant space of approximants, developing also preconditioning schemes [2].
Recently the two papers [11, 12] introduced a new tool, a general way to produce stable bases for the native space $\mathcal{N}_\Phi(\Omega)$ associated with a kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$, based on a suitable factorization of the kernel matrix $A := (\Phi(x_i, x_j))$. In [11] this process was used to create a *Newton basis*, which turns out to be stable, complete, orthonormal and recursively computable. A related approach, based on a different matrix decomposition, is used also in [6].

We use these ideas and we refine them in order to build a *new stable basis* that shares some properties with the one proposed in [12]. Moreover, the new basis provides a connection with a “natural” basis for the functional space $\mathcal{N}_\Phi(\Omega)$, which arises from an eigendecomposition of a compact integral operator associated with the kernel Φ, and which brings intrinsic information about the kernel Φ and the set Ω. In the recent paper [8] this eigenbasis in the case of the Gaussian kernel was also used, altough in a different way, to produce fast and stable interpolants and least-squares approximants.
From the numerical point of view, the structure of the new basis allows to further stabilize the approximation by moving from an exact data interpolation to an approximation in the least-squares sense, with a process that exactly corresponds to a low-rank approximation of the kernel matrix.

The paper is organized as follows. In Section 2 we introduce the so-called eigenbasis while in Section 3 we describe the techniques introduced in [11, 12], which will be our starting point for the development of the new basis, which is constructed and analyzed in Section 4. Section 5 presents some numerical examples which test the new basis on various domains and kernels. The last Section discusses future works that could be done for improving and better understanding our results.

2. An integral operator and a ”natural” basis

Given a set $\Omega \subset \mathbb{R}^s$ and a positive definite kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$ such that, for all $x, y \in \Omega$, $\Phi(x, y) = \phi(\|x - y\|_2)$, where $\phi : [0, \infty) \to \mathbb{R}$, we can define an integral operator T_Φ associated to the kernel Φ. The construction and the properties of this operator are discussed in detail in [15, §10.4]. Here we are mainly interested in a particular basis that arises from an eigendecomposition of it.
Consider the operator $T_\Phi : L_2(\Omega) \to \mathcal{N}_\Phi(\Omega) \subset L_2(\Omega)$ defined by

$$T_\Phi[f](x) := \int_\Omega \Phi(x, y) f(y) dy \quad \forall f \in L_2(\Omega),\ \forall x \in \Omega\,, \tag{1}$$

that maps $L_2(\Omega)$ continuously into $\mathcal{N}_\Phi(\Omega)$. It is the adjoint of the embedding operator of $\mathcal{N}_\Phi(\Omega)$ into $L_2(\Omega)$, i.e.

$$(f, v)_{L_2(\Omega)} = (f, T_\Phi[v])_\Phi \quad \forall f \in \mathcal{N}_\Phi(\Omega),\ \forall v \in L_2(\Omega)\,. \tag{2}$$

A particular and in some sense “natural“ basis for $\mathcal{N}_\Phi(\Omega)$ comes from the famous *Mercer’s theorem* (cf. [10]).

Theorem 1. *Every continuous positive definite kernel Φ on a bounded domain $\Omega \subset \mathbb{R}^s$ defines an operator*

$$T_\Phi : \mathcal{N}_\Phi(\Omega) \to \mathcal{N}_\Phi(\Omega), \quad T_\Phi[f] = \int_\Omega \Phi(x, y) f(y) dy\,,$$

which is bounded, compact and self-adjoint. It has an enumerable set of eigenvalues $\{\lambda_j\}_{j>0}$ and eigenvectors $\{\varphi_j\}_{j>0}$, i.e. $\forall j > 0$

$$\lambda_j \varphi_j(x) = \int_\Omega \Phi(x, y) \varphi_j(y) dy \quad \forall x \in \Omega\,,$$

which form an orthonormal basis for $\mathcal{N}_\Phi(\Omega)$. In particular

$$\begin{aligned} &\{\varphi_j\}_{j>0} && \textit{is orthonormal in } \mathcal{N}_\Phi(\Omega)\,, \\ &\{\varphi_j\}_{j>0} && \textit{is orthogonal in } L_2(\Omega),\ \|\varphi_j\|^2_{L_2(\Omega)} = \lambda_j\,, \\ &\lambda_j \to 0 && \textit{as } j \to \infty\,. \end{aligned}$$

Moreover, the kernel has a series expansion

$$\Phi(x,y) = \sum_{j=1}^{\infty} \varphi_j(x)\varphi_j(y) \quad \forall x,y \in \Omega\,,$$

where the $\{\varphi_j\}$ are eigenfunctions, which is absolutely and uniformly convergent.

Remark 2. The operator T_Φ is a trace-class operator, that is to say

$$\sum_{j>0} \lambda_j = \int_\Omega \Phi(x,x)\ dx = \phi(0)\ |\Omega|\,,$$

where $|\Omega| := \text{meas}(\Omega)$.

This property, together with the fact that the eigenvalues accumulate in 0, will be useful to estimate the convergence of the truncated series with respect to the whole one. Moreover, as a consequence of the Property (2), which we point out for later use, for all $j > 0$

$$(f,\varphi_j)_{L_2(\Omega)} = (f, T_\Phi[\varphi_j])_\Phi = \lambda_j\ (f,\varphi_j)_\Phi = (\varphi_j,\varphi_j)_{L_2(\Omega)}\ (f,\varphi_j)_\Phi \quad \forall f \in \mathcal{N}_\Phi(\Omega)\,.$$

3. General bases

In this section we give a brief account of the tools introduced in [11, 12]. Among the results discussed in details in those papers, we stress the connection between a change of basis and a decomposition of the kernel matrix A, together with a characterization of such bases.

Let $\Omega \subset \mathbb{R}^s$, $X = \{x_1, \ldots, x_N\} \subset \Omega$ and let $\mathcal{T}_X = \{\Phi(\cdot, x_i), x_i \in X\}$ be the standard basis of translates. Consider another basis $\mathcal{U} = \{u_i \in\ \mathcal{N}_\Phi(\Omega),\ i = 1,\ldots,N\}$ such that

$$\text{span}(\mathcal{U}) = \text{span}(\mathcal{T}_X) =: \mathcal{N}_\Phi(X)\,. \tag{3}$$

At $x \in \Omega$, $\mathcal{T}_X$ and $\mathcal{U}$ can be expressed as the row vectors

$$\begin{aligned} T(x) &= [\Phi(x,x_1),\ldots,\Phi(x,x_N)] \in \mathbb{R}^N\,, \\ U(x) &= [u_1(x),\ldots,u_N(x)] \in \mathbb{R}^N\,. \end{aligned}$$

The following theorem gives a characterization of the basis $\mathcal{U}$.

Theorem 3. *Any basis $\mathcal{U}$ arises from a factorization of the kernel matrix A, i.e.*

$$A = V_\mathcal{U} \cdot {C_\mathcal{U}}^{-1}\,,$$

where $V_\mathcal{U} = (u_j(x_i))_{1\leqslant i,j\leqslant N}$ and the coefficient matrix $C_\mathcal{U}$ is such that $U(x) = T(x)\cdot C_\mathcal{U}$.

This factorization allows to express the interpolant of a given function $f \in \mathcal{N}_\Phi(\Omega)$ in the following way.

Proposition 4. *The interpolant $P_X[f]$ on $X \subset \Omega$ of a function $f \in \mathcal{N}_\Phi(\Omega)$ can be rewritten as*

$$P_X[f](x) = \sum_{j=1}^{N} \Lambda_j(f)\ u_j(x) = U(x)\cdot\Lambda(f) \quad \forall x \in \Omega\,,$$

where $\Lambda(f) = [\Lambda_1(f),\ldots,\Lambda_N(f)]^T \in \mathbb{R}^N$ is a column vector of values of linear functionals defined by

$$\Lambda(f) = {C_\mathcal{U}}^{-1}\cdot A^{-1}\cdot E_X(f) = {V_\mathcal{U}}^{-1}\cdot E_X(f)\,,$$

while $E_X(f)$ is the column vector given by the evaluations of f on X.

Now we can give a *stability estimate*, which considers the particular basis used.

Proposition 5. *Let* $G_{\mathcal{U}} := ((u_i, u_j)_\Phi)_{1 \leqslant i,j \leqslant N}$, $\rho(G_{\mathcal{U}})$ *the spectral radius and* $\kappa_2(G_{\mathcal{U}})$ *the corresponding 2-condition number. Then,* $\forall x \in \Omega$,

$$|P_X[f](x)|^2 \leqslant \|U(x)\|_2^2 \; \|\Lambda_{\mathcal{U}}(f)\|_2^2 \leqslant \kappa_2(G_{\mathcal{U}}) \; \phi(0) \; \|f\|_\Phi^2 . \tag{4}$$

In particular

$$\begin{array}{rcl} \|U(x)\|_2^2 & \leqslant & \rho(G_{\mathcal{U}}) \; \phi(0) \quad \forall x \in \Omega \,, \\ \|\Lambda_{\mathcal{U}}(f)\|_2^2 & \leqslant & \rho(G_{\mathcal{U}}^{-1}) \; \|f\|_\Phi^2 \quad \forall f \in \mathcal{N}_\Phi(\Omega) \,. \end{array}$$

The message contained in this Proposition is that orthonormal bases have to be considered for stability purpose. Indeed, the next corollary shows that they give the best results in terms of stability.

Corollary 6. *If* $\mathcal{U}$ *is a* Φ*-orthonormal basis, the stability estimate (4) becomes*

$$|P_X[f](x)| \leqslant \sqrt{\phi(0)} \; \|f\|_\Phi \quad \forall x \in \Omega \,. \tag{5}$$

In particular, for fixed $x \in \Omega$ *and* $f \in \mathcal{N}_\Phi(\Omega)$, *the values* $\|U(x)\|_2$ *and* $\|\Lambda(f)\|_2$ *are the same for all* Φ*-orthonormal bases independently on* X

$$\|U(x)\|_2 \leqslant \sqrt{\phi(0)}, \quad \|\Lambda(f)\|_2 \leqslant \|f\|_\Phi \,.$$

To conclude this section, we recall that in order to build such a basis it is enough to choose a particular matrix decomposition. This is the idea on which relies the construction of the new basis.

Theorem 7. *Each* Φ*-orthonormal basis* $\mathcal{U}$ *arises from a decomposition*

$$A = B^T \cdot B,$$

with $B^T = V_{\mathcal{U}}$ *and* $B^{-1} = C_{\mathcal{U}}$.

A similar result holds for $\ell_2(X)$-orthonormal bases.

Theorem 8. *Each* $\ell_2(X)$*-orthonormal basis* $\mathcal{U}$ *arises from a decomposition*

$$A = Q \cdot B \,,$$

with Q *orthonormal,* $Q = V_{\mathcal{U}}$ *and* $B = C_{\mathcal{U}}$.

4. Weighted SVD bases

The main idea for the construction of the new basis is to discretize the "natural" basis introduced in Theorem 1. To this aim, consider on Ω a cubature rule $(X, \mathcal{W})_N$, $N \in \mathbb{N}$, that is a set of distinct points $X = \{x_j\}_{j=1}^N \subset \Omega$ and a set of *positive* weights $\mathcal{W} = \{w_j\}_{j=1}^N$ such that

$$\int_\Omega f(y)dy \approx \sum_{j=1}^N f(x_j) w_j \quad \forall f \in \mathcal{N}_\Phi(\Omega) \,.$$

This allows to approximate the operator (1) for each eigenvalue λ_j, $j > 0$, using the *symmetric Nyström method* (c.f. e.g [1, §11.4]) based on the above cubature rule.
Thus, the operator T_Φ can be evaluated on X as

$$\lambda_j \varphi_j(x_i) = \int_\Omega \Phi(x_i, y)\varphi_j(y)dy \quad i = 1, \ldots, N, \ \forall j > 0,$$

and then discretized using the cubature rule by

$$\lambda_j \varphi_j(x_i) \approx \sum_{h=1}^{N} \Phi(x_i, x_h)\varphi_j(x_h)w_h \quad i, j = 1, \ldots, N. \tag{6}$$

Now, setting $W = \mathrm{diag}(w_j)$, it suffices to solve the following discrete eigenvalue problem in order to find the approximation of the eigenvalues and eigenfunctions (evaluated on X) of $T_\Phi[f]$:

$$\lambda v = (A \cdot W)v.$$

This approach does not lead directly to the connection between the discretized version of the basis of Theorem 1 and a basis of the subspace $\mathcal{N}_\Phi(X)$. In fact it involves a scaled version of the kernel matrix, that is $A \cdot W$, which is no longer symmetric and that cannot be described as a factorization of A, as required by the construction made in the previous section.
A solution is to rewrite (6) using the fact that the weights are positive as

$$\lambda_j(\sqrt{w_i}\varphi_j(x_i)) = \sum_{h=1}^{N} (\sqrt{w_i}\Phi(x_i, x_h)\sqrt{w_h})(\sqrt{w_h}\varphi_j(x_h)) \quad \forall i, j = 1, \ldots, N\,,$$

and then to consider the corresponding scaled eigenvalue problem

$$\lambda \left(\sqrt{W} \cdot v\right) = \left(\sqrt{W} \cdot A \cdot \sqrt{W}\right) \left(\sqrt{W} \cdot v\right)$$

which is equivalent to the previous one, now involving the symmetric and positive definite matrix $A_W := \sqrt{W} \cdot A \cdot \sqrt{W}$. This matrix is normal, then a *singular value decomposition* of A_W is a *unitary diagonalization.*

Motivated by this approach we can introduce a *weighted SVD basis* for $\mathcal{N}_\Phi(X)$, described in terms of the notation given in Theorem 3.

Definition 9. *A weighted SVD basis $\mathcal{U}$ is a basis for $\mathcal{N}_\Phi(X)$ characterized by the following matrices:*

$$V_\mathcal{U} = \sqrt{W^{-1}} \cdot Q \cdot \Sigma, \quad C_\mathcal{U} = \sqrt{W} \cdot Q \cdot \Sigma^{-1}$$

where

$$\sqrt{W} \cdot A \cdot \sqrt{W} = Q \cdot \Sigma^2 \cdot Q^T$$

is a singular value decomposition (and a unitary diagonalization) of the scaled kernel matrix A_W. To be more precise, Σ is a diagonal matrix with $\Sigma_{jj} = \sigma_j$, $j = 1, \ldots, N$ and $\sigma_1^2 \geqslant \cdots \geqslant \sigma_N^2 > 0$ are the singular values of A_W, and W is a diagonal matrix where $W_{jj} = w_j$, $j = 1, \ldots, N$ are the weigths of the cubature rule $(X, \ \mathcal{W})_N$.

In what follows it is of fundamental importance to require that $\displaystyle\sum_{j=1}^{N} w_j = |\Omega|$, which is equivalent to ask that for all $N \in \mathbb{N}$ the cubature rule $(X, \ \mathcal{W})_N$ is exact at least for the constant functions.

As expected, the weighted SVD basis preserves some interesting properties of the "natural" one, as stated in the next Theorem. From now on, we denote as $\ell_{2,w}(X)$ the $\ell_2(X)$ inner product weighted with the weights of the cubature rule $(X,\ \mathcal{W})_N$, which are assumed to be positive.

Theorem 10. *Every weighted SVD basis $\mathcal{U}$ satisfies:*

1. $u_j(x) = \dfrac{1}{\sigma_j^2}\displaystyle\sum_{i=1}^{N} w_i u_j(x_i)\Phi(x, x_i) \approx \dfrac{1}{\sigma_j^2} T_\Phi[u_j](x),\ \forall\ 1 \leqslant j \leqslant N,\ \forall x \in \Omega;$
2. *$\mathcal{U}$ is Φ-orthonormal;*
3. *$\mathcal{U}$ is $\ell_{2,w}(X)$-orthogonal;*
4. $\|u_j\|^2_{\ell_{2,w}(X)} = \sigma_j^2 \quad \forall u_j \in \mathcal{U};$
5. $\sum_{j=1}^{N} \sigma_j^2 = \phi(0)\ |\Omega|.$

Proof: Properties 2. - 3. and 4. can be proved using the expression for the gramians computed in [12], where in this case the $\ell_2(X)$ products are weighted in the proper way. Indeed,

$$
\begin{aligned}
G_{\mathcal{U}} &= {C_{\mathcal{U}}}^T \cdot A \cdot C_{\mathcal{U}} = {C_{\mathcal{U}}}^T \cdot V_{\mathcal{U}} = \Sigma^{-1} \cdot Q^T \cdot \sqrt{W} \cdot \sqrt{W^{-1}} \cdot Q \cdot \Sigma = I \\
\Gamma_{\mathcal{U}} &= {V_{\mathcal{U}}}^T \cdot W \cdot V_{\mathcal{U}} = \Sigma \cdot Q^T \cdot \sqrt{W^{-1}} \cdot W \cdot \sqrt{W^{-1}} \cdot Q \cdot \Sigma = \Sigma^2
\end{aligned}
$$

To prove Property 1. it suffices to use the definition of $C_{\mathcal{U}}$ and $V_{\mathcal{U}}$. Indeed, from the definition of $V_{\mathcal{U}}$, denoting by $V_{\mathcal{U}j}$ the j-th column of $V_{\mathcal{U}}$, we get

$$
\begin{aligned}
V_{\mathcal{U}} &= \sqrt{W^{-1}} Q \Sigma = \sqrt{W^{-1}}[q_1\sigma_1, \ldots, q_N\sigma_N] \\
\Rightarrow &\qquad E_X(u_j) = V_{\mathcal{U}j} = (\sqrt{W})^{-1} q_j \sigma_j \\
\Rightarrow &\qquad q_j/\sigma_j = 1/\sigma_j^2 \sqrt{W} E_X(u_j)\,.
\end{aligned}
$$

Using the last equality we can compute each component of $C_{\mathcal{U}}$ as

$$
(C_{\mathcal{U}})_{i,j} = (\sqrt{W} \cdot Q \cdot \Sigma^{-1})_{i,j} = \sqrt{w_i}\ \frac{q_j(i)}{\sigma_j} = \frac{w_i}{\sigma_j^2}\ u_j(x_i)
$$

and then by the Definition of $\mathcal{U}$

$$
\begin{aligned}
u_j(x) &= \sum_{i=1}^{N} \Phi(x, x_i)\ (C_{\mathcal{U}})_{i,j} = \sum_{i=1}^{N} \Phi(x, x_i)\ \frac{w_i}{\sigma_j^2}\ u_j(x_i) = \\
&= \frac{1}{\sigma_j^2} \sum_{i=1}^{N} w_i\ \Phi(x, x_i)\ u_j(x_i)\,,
\end{aligned}
$$

where the last term is clearly the approximation of $T_\Phi[u_j]$ given by the cubature rule $(X,\ \mathcal{W})_N$, divided by the corresponding discrete eigenvalue σ_j^2.
Finally the Property 5. is proved by linear algebra arguments. Recalling that

$$
\sqrt{W} \cdot A \cdot \sqrt{W} = Q \cdot \Sigma^2 \cdot Q^T\,,
$$

and the fact that the trace of a square matrix is equal to the sum of its eigenvalues, we get

$$
\sum_{j=1}^{N} \sigma_j^2 = \sum_{j=1}^{N} w_j\ \Phi(x_j, x_j) = \phi(0)\ \sum_{j=1}^{N} w_j = \phi(0)\ |\Omega|\,.
$$

This concludes the proof. $\square$

It is important to notice that in this context, when $\{w_j\}_{j=1}^N$ are cubature weights, the $\ell_{2,w}(X)$-scalar product is a discretization of the $L_2(\Omega)$-scalar product. Indeed

$$(f,g)^2_{L_2(\Omega)} = \int_\Omega f(x)g(x)dx \approx \sum_{j=1}^N w_j f(x_j)g(x_j) = (f,g)^2_{\ell_{2,w}(X)} \,.$$

Using this, the Property 3. is simply a discretized version of the corresponding property of the continuous basis. Moreover, Property 5. suggests that

$$\sum_{j=1}^N \sigma_j^2 = \sum_{j=1}^N w_j \;\Phi(x_j,x_j) = \int_\Omega \Phi(x,x)\; dx \,,$$

which is exactly the relation of the continuous eigenvalues if $N \to \infty$, as pointed out in the Remark 2. In fact, in this case the integral is exactly approximated by the cubature rule (it is supposed to be exact at least for constant functions).
We point out that to construct the basis we require just that the weigths $\{w_i\}_{i=1}^N$ are positive and able to reproduce constants. In principle it is possible to use weights not related to a cubature rule, but in this way no connection can be expected between $\mathcal{U}$ and the eigenbasis $\{\varphi_j\}_{j>0}$, while remains unchanged the stabilization properties due to the use of a Singular Value Decomposition, possibly truncated.

Moreover, the basis has some further properties due to its orthonormality in $\mathcal{N}_\Phi(\Omega)$, as pointed out again in [12]. Indeed, it is a complete basis for the full space if the data set X is dense in Ω, and the norm of the pointwise-error operator, namely the *power function* $\mathcal{P}_{\Phi,X}(x)$, can be expressed in the simpler form

$$\mathcal{P}_{\Phi,X}(x)^2 = \phi(0) - \sum_{j=1}^N u_j(x)^2 \,, \tag{7}$$

which involves only the basis under consideration. Furthermore, it is possible to give an expansion of the kernel when it acts on functions of $\mathcal{N}_\Phi(X)$. That is

$$\Phi(x,y) = \sum_{j=1}^N u_j(x)u_j(y) \quad \forall x,y \in \Omega.$$

This fact is useful when it is required to use a degenerate kernel to approximate the original one.

4.1. Weighted discrete least-squares approximation

Now we can introduce a weighted discrete least-squares operator that turns out to be strictly related to the weighted basis just introduced. The goal is to project the unknown function $f \in \mathcal{N}_\Phi(\Omega)$ into a proper subset of $\mathcal{N}_\Phi(X)$. In other words we will use a smaller basis $\mathcal{U}' \subsetneq \mathcal{U}$. This is done in order to obtain better results in terms of stability and computational cost, without serious loss of convergence speed. This kind of approximation is meaningful when the data values are supposed to be affected by noise, or when the kernel matrix A is seriously ill-conditioned.
The weighted least-squares approximation we are interested in can be so defined.

Definition 11. *Given a function $f \in \mathcal{N}_\Phi(\Omega)$, a discrete subset $X \subset \Omega$, a set of cubature weights $\mathcal{W}$ associated with X, a weighted SVD basis $\mathcal{U}$ for $\mathcal{N}_\Phi(X)$ and a natural number $M \leqslant N = |X|$, the weighted discrete least-squares approximation of order M of f is the function $\Lambda_M[f]$ that satisfies the condition*

$$\Lambda_M[f] = \underset{g \in \mathrm{span}\{u_1,\ldots,u_M\}}{\arg\min} \|f - g\|_{\ell_{2,w}(X)} \, .$$

We need a relation between the Φ and $\ell_{2,w}(X)$-inner products. The last is again a discretized version of a property of the continuous basis $\{\varphi_j\}_{j>0}$, stated in Remark 2.

Lemma 12. *For all $f \in \mathcal{N}_\Phi(\Omega)$ and for each $u_j \in \mathcal{U}$, the following relation between the Φ and $\ell_{2,w}(X)$-inner products holds:*

$$(f, u_j)_\Phi = \frac{1}{\sigma_j^2} (f, u_j)_{\ell_{2,w}(X)} = \frac{(f, u_j)_{\ell_{2,w}(X)}}{(u_j, u_j)_{\ell_{2,w}(X)}} \, .$$

Proof: Using the Property 1. of Theorem 10 and by direct calculations we get:

$$\begin{aligned} (f, u_j)_\Phi &= \left(f, \frac{1}{\sigma_j^2} \sum_{i=1}^N w_i u_j(x_i) \Phi(\cdot, x_i) \right)_\Phi = \frac{1}{\sigma_j^2} \sum_{i=1}^N w_i u_j(x_i) \left(f, \Phi(\cdot, x_i)\right)_\Phi \, , \\ &= \frac{1}{\sigma_j^2} \sum_{i=1}^N w_i u_j(x_i) f(x_i) = \frac{1}{\sigma_j^2} (f, u_j)_{\ell_{2,w}(X)} \, , \end{aligned}$$

where, by the Property 4. of Proposition 10, $\sigma_j^2 = (u_j, u_j)_{\ell_{2,w}(X)}$. $\Box$

Using this Lemma with the notation of Definition 11, it comes easy to compute the weighted discrete least-squares approximation of a function $f \in \mathcal{N}_\Phi(\Omega)$.

Theorem 13. *The weighted discrete least-squares approximation of a function $f \in \mathcal{N}_\Phi(\Omega)$ is given by*

$$\Lambda_M[f](x) = \sum_{j=1}^M \frac{(f, u_j)_{\ell_{2,w}(X)}}{\sigma_j^2} u_j(x) = \sum_{j=1}^M (f, u_j)_\Phi u_j(x) \, , \quad \forall x \in \Omega \, , \tag{8}$$

that is $\Lambda_M[f]$ is nothing else but a truncation to the first M terms of $P_X[f]$.

Remark 14. We observe that to compute the weighted least-squares approximant $\Lambda_M[f]$ it suffices to use the first M elements of the interpolant $P_X[f]$, which correspond to the biggest singular values σ_j^2, $j = 1, \ldots, N$. This is opposite to the case of the standard basis of translates, where the choise of the elements of the basis to neglect correspond to the choice of a restricted subset $Y \subset X$.
The idea of constructing this weighted discrete least-squares approximant, can be automated when dealing with very small singular values. In this case, in order to avoid numerical instability, we can leave out the basis corresponding to singular values less than a pre-assigned tolerance, skipping automatically from interpolation to discrete least-squares approximation. From a linear algebra point of view, this corresponds to solve the (weighted) linear system associated to the interpolation problem using a *total least-squares* method.

Remark 15. As a result of using an orthonormal basis, the pointwise-evaluation stability can be bounded using the general estimates introduced in the previous section (relation (4)). We also stress that the norm of the pointwise-evaluation operator $\left\|\mathcal{E}_x^M\right\|_{\mathcal{N}_\Phi(\Omega)^*}$ related to the approximation operator, which is the equivalent to the power function, is simply given by

$$\left\|\mathcal{E}_x^M\right\|^2_{\mathcal{N}_\Phi(\Omega)^*} = \phi(0) - \sum_{j=1}^{M} u_j(x)^2\,.$$

It is clear that it is a simple truncation of the power function. Altough it is an obvious consequence of Theorem 13, there is no reason in general to expect such a relation if we consider a different kind of least-squares approximant in an arbitrary basis. Moreover, it is also clear that by replacing an exact interpolant with a weighted discrete least-squares approximant, we can obtain better results in terms of stability. In particular, the stability estimate (4) can be also refined for the particular case of a weighted SVD-basis, as stated in the next Corollary.

Corollary 16. *For a weighted SVD basis $\mathcal{U}$ and a function $f \in \mathcal{N}_\Phi(\Omega)$, the following stability estimate holds:*

$$|\Lambda_M[f](x)| \leqslant \sqrt{\phi(0)}\;\|f\|_\Phi\,, \quad \forall x \in \Omega$$

In particular,

$$|\Lambda_M[f](x)| \leqslant \sqrt{\sum_{j=1}^{M} u_j(x)^2}\;\|f\|_\Phi\,, \quad \forall x \in \Omega\,.$$

Proof: The result is a direct consequence of the Cauchy-Schwartz inequality applied to the second equality in (8). □

4.2. Error bounds

We can now prove some convergence estimates for these approximants. It is important to notice that, in the case of interpolation, there is no difference in using a particular basis since the spanned subspace $\mathcal{N}_\Phi(X)$ defined in (3), in which we project a function $f \in \mathcal{N}_\Phi(\Omega)$, clearly does not depend on the chosen basis.
On the other hand, the fact that we are using this kind of basis allows us to connect the bounds to the continuous eigenvalues and to their eigenfunctions $\{\varphi_j\}_{j>0}$, which are related in a close way to the kernel Φ. This justifies the choice of sampling the function f on the data set X that, together with the weights $\mathcal{W}$, form a good cubature rule.
This observation remains valid in the case of the weighted least-squares approximant for which, in addition, the connection between the discrete and the continuous eigenvalues motivates the use of a reduced subspace of $\mathcal{N}_\Phi(X)$. This is confirmed by the following theorem.

Theorem 17. *Let $\Omega \subset \mathbb{R}^s$ be compact, let $\Phi \in \mathcal{C}(\Omega \times \Omega)$ be a radial positive definite kernel and $X \subset \Omega$. Then, there exist a constant C, depending on Φ, Ω, X and $\mathcal{W}$, such that for each $f \in \mathcal{N}_\Phi(\Omega)$*

$$\|f - P_X[f]\|^2_{L_2(\Omega)} \leqslant \left(|\Omega|\cdot\phi(0) - \sum_{j=1}^{N}\lambda_j + C\cdot\sum_{j=1}^{N}\|u_j - \varphi_j\|_{L_2(\Omega)}\right)\|f\|^2_\Phi\,.$$

Proof: Using the expression (7) for the power function, we can write the standard error bound for interpolation as

$$|f(x) - P_X[f](x)|^2 \leqslant \left(\phi(0) - \sum_{j=1}^{N} u_j(x)^2\right)\;\|f\|^2_\Phi\,.$$

From the embedding $\mathcal{N}_\Phi(\Omega) \hookrightarrow L_2(\Omega)$ we know that both sides have finite $L_2(\Omega)$-norm. Thus we can integrate over Ω the bound and get

$$\begin{aligned}
\|f - P_X[f]\|^2_{L_2(\Omega)} &\leqslant \int_\Omega \left(\phi(0) - \sum_{j=1}^N u_j(x)^2\right) \|f\|^2_\Phi dx\,, \\
&= \left(|\Omega| \cdot \phi(0) - \sum_{j=1}^N \int_\Omega u_j(x)^2 dx\right) \|f\|^2_\Phi \\
&= \left(|\Omega| \cdot \phi(0) - \sum_{j=1}^N \|u_j(x)\|^2_{L_2(\Omega)}\right) \|f\|^2_\Phi\,.
\end{aligned}$$

Using the relations

$$\begin{aligned}
\|\varphi_j\|^2_{L_2(\Omega)} &= \|u_j\|^2_{L_2(\Omega)} + \|\varphi_j - u_j\|^2_{L_2(\Omega)} + 2\ (\varphi_j - u_j, u_j)_{L_2(\Omega)} \\
\|u_j\|_{L_2(\Omega)} &\leqslant \|\varphi_j - u_j\|_{L_2(\Omega)} + \|\varphi_j\|_{L_2(\Omega)}
\end{aligned}$$

then for all $j = 1, \ldots, N$ we can estimate the $L_2(\Omega)$-norm as

$$\begin{aligned}
-\|u_j\|^2_{L_2(\Omega)} &= -\|\varphi_j\|^2_{L_2(\Omega)} + \|\varphi_j - u_j\|^2_{L_2(\Omega)} + 2\ (\varphi_j - u_j, u_j)_{L_2(\Omega)} \\
&\leqslant -\|\varphi_j\|^2_{L_2(\Omega)} + \|\varphi_j - u_j\|^2_{L_2(\Omega)} + 2\ \|\varphi_j - u_j\|_{L_2(\Omega)}\ \|u_j\|_{L_2(\Omega)} \\
&\leqslant -\|\varphi_j\|^2_{L_2(\Omega)} + \|\varphi_j - u_j\|_{L_2(\Omega)} \left(3\ \|\varphi_j - u_j\|_{L_2(\Omega)} + \|\varphi_j\|_{L_2(\Omega)}\right)\,.
\end{aligned}$$

From (1) we know that $\|\varphi_j\|^2_{L_2(\Omega)} = \lambda_j\ \forall j > 0$. Then, to conclude we observe that

$$3\ \|\varphi_j - u_j\|_{L_2(\Omega)} + \|\varphi_j\|_{L_2(\Omega)} \leqslant 3\left(\max_{j=1,\ldots,N} \|\varphi_j - u_j\|_{L_2(\Omega)}\right) + \sqrt{\lambda_1} =: C$$

since the eigenvalues $\{\lambda_j\}_{j>0}$ are not increasing. □

The same estimate remains valid for the approximant $\Lambda_M[f]$ if N is replaced by M, as a consequence of the Remark 15.

Remark 18. We point out that these error estimates involve two terms.

- The first one, $|\Omega| \cdot \phi(0) - \sum_{j=1}^N \lambda_j$, is related only to the kernel Φ, the domain Ω and the dimension $N \in \mathbb{N}$ of the approximation subspace $\mathcal{N}_\Phi(X)$. From the Remark 2 of the Theorem 1 we know that for $N \to \infty$ the above term vanishes, and the eigenvalues are positive and orderer in decreasing way. Hence, this term measures how the truncated series approximates the full one, or in other words, how the degenerate kernel

$$\sum_{j=1}^N \lambda_j \varphi_j(x) \varphi_j(y),\ \ x, y \in \Omega$$

  approximates the original kernel $\Phi(x, y)$.

- The second term, $C \cdot \sum_{j=1}^N \|u_j - \varphi_j\|_{L_2(\Omega)}$, depends on the cubature rule $(X,\ \mathcal{W})_N$. It measures the convergence rate of the Nystöm method based on the rule, and gives information on how well the discrete basis $\mathcal{U}$ approximates the continuous one.

In this view, it is important to notice what is the effect of an high-order cubature formula. Obviously a higher accuracy will lead to a better approximation of the eigenbasis $\{\varphi_j\}_{j>0}$, so in the limit we could expect that our basis will be able to reproduce $\{\varphi_j\}_{j>0}$ and each function in $\mathcal{N}_\Phi(\Omega)$ (hopefully in a good way). Nevertheless, at a finite stage with fixed N, also assuming to know an almost exact cubature formula, we are still approximating $f \in \mathcal{N}_\Phi(\Omega)$ with a projection into $\mathcal{N}_\Phi(X)$. This means that even if the second term is reduced by a suitable choice of $(X,\ \mathcal{W})_N$, the first one still depends in a strict way on other parameters of the approximation process.

For the weighted least-squares approximant, it make sense to consider also another type of error. Indeed in this case the data set $X \subset \Omega$ is not used to interpolate the function f, but only as a set of samples. So, the pointwise error between f and $\Lambda_M[f]$ on X is not zero. We can bound this quantity as well as shown in the next Proposition.

Proposition 19. *Let $\Omega \subset \mathbb{R}^s$, $\Phi \in \mathcal{C}(\Omega \times \Omega)$ a radial positive definite kernel, $X \subset \Omega$, $|X| = N$, and $M < N$. Then, for every $f \in \mathcal{N}_\Phi(\Omega)$,*

$$\|f - \Lambda_M[f]\|_{\ell_{2,w}(X)} \leqslant \left(\sum_{j=M+1}^{N} \sigma_j^2 \right)^{\frac{1}{2}} \|f\|_\Phi \tag{9}$$

This proposition can be proved as we did in the previous Theorem, replacing in the proper way the $L_2(\Omega)$ norm with the discrete $\ell_{2,w}(X)$ norm.

This result can be interpreted also in another way. In fact, it gives a bound on how much the weighted least-squares approximant and the interpolant differ on the set X. Indeed, since $f(x_i) = P_X[f](x_i)\ \forall x_i \in X$, we can replace f with $P_X[f]$ in the left-hand side of estimate (9). The well-known *trade-off principle* between stability and convergence in approximations by radial basis functions (see e.g. [15, §12.1]) can be viewed in this context as follows. We have $\forall f \in \mathcal{N}_\Phi(\Omega)$

$$\begin{aligned} |P_X[f](x)|^2 &\leqslant \left(\sum_{j=1}^{N} u_j(x)^2 \right) \|f\|_\phi^2 \,, \\ |P_X[f](x) - f(x)|^2 &\leqslant \left(\phi(0) - \sum_{j=1}^{N} u_j(x)^2 \right) \|f\|_\phi^2 \,, \end{aligned}$$

and the same for $\Lambda_M[f]$ if N is replaced by $M < N$.
Hence, for convergent approximations, namely for approximations for which the power function decreases to zero, we have necessarily

$$\sum_{j=1}^{N} u_j(x)^2 \to \phi(0)\,.$$

That is, the stability bound is maximized by $\phi(0)$, in accordance with the Proposition 16.

5. Numerical examples

For testing the behavior of the basis we need a set $X \subset \Omega$ of data sites such that $(X,\ \mathcal{W})_N$ is a cubature rule. Suitable sets are for example

- the *trigonometric gaussian points*, recently studied and computed in [3, 14], which can be obtained for a wide class of domains, such as the disk, circular zones and circular lenses;

- the product Gauss-Legendre well-known for the cubature on the cube.

Concerning the trigonometric gaussian points, it is worth mentioning that they provide high-accuracy cubature rules while being sufficiently uniformly distributed in Ω. As a computional issue, *Matlab* functions to compute these points can be found in the web site [16]. The code is mainly *Matlab*, using in some parts the software from the book [7]. Only some parts, the most critical from the performance point of view, are written in *C++*, using the *MatlabMEX* interface (cf. [17]) and the linear algebra package *Eigen* (see [18]).

For further examples we refer the reader to the Master's thesis [13, Ch. 4], where we compare also the new basis with the Newton basis introduced in [11]. We observe that the Newton basis is built by a recursive and adaptive choice of the data sites X. This capability makes difficult a comparison with the new basis, since it requires to fix a cubature set $(X,\ \mathcal{W})_N$.

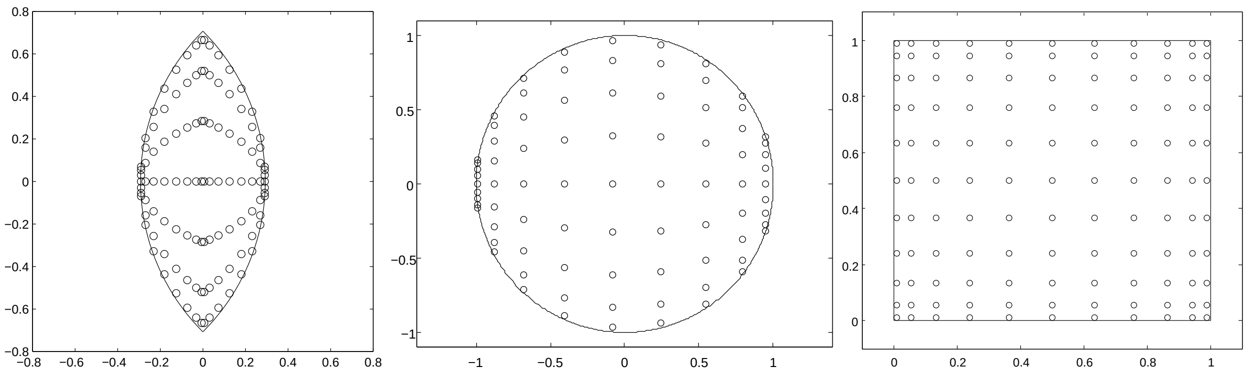

Figure 1: The domains used in the numerical experiments with an example of the corresponding sample points. From left to right: the lens Ω_1 (trigonometric-gaussian points), the disk Ω_2 (trigonometric-gaussian points) and the square Ω_3 (product Gauss-Legendre points).

5.1. Comparison between the interpolant and the weighted least-squares approximant

As pointed out in the Remark 14, we know that we can compute the weighted least-squares approximant as the truncation of the interpolant. This approach increases the residual, as shown in Proposition 17 and in the subsequent discussion, but in the cases in which the smaller eigenvalues are under a certain tolerance, we may expect that a truncation does not affect too much the approximation properties of the process. Although in Corollary 16 we proved the stability of the new basis, in some limit situations we could expect that the influence of the smallest eigenvalues produces numerical instability that could not completely be controlled.
Here we compare the approximation error produced using the full interpolant and some reduced weighted least-squares approximant starting from 600 trigonometric gaussian centers, and then truncating the basis for $M \in \{0, 20, \dots, 600\}$.
We reconstruct the oscillatory function $f_o(x, y) = \cos(20(x + y))$ on the disk Ω_1 with center $C = \left(\frac{1}{2}, \frac{1}{2}\right)$ and radius $R = \frac{1}{2}$. The experiment has been repeated for three different kernels: the gaussian kernel, the inverse multiquadric (IMQ) and the cubic Matérn kernel (3MAT), which is generated by the univariate function

$$\phi(r) = e^{-\varepsilon r}(15 + 15\varepsilon r + 6(\varepsilon r)^2 + (\varepsilon r)^3),\ r \geqslant 0$$

for a shape parameter $\varepsilon = 1, 4, 9$. This choice is motivated from the different behavior of the eigenvalues $\{\lambda_j\}_{j>0}$ of the operator T_Φ associated with these radial basis functions. Indeed, although we know from Theorem 1 that the continuous eigenvalues accumulate to zero, the speed to which

they decay is clearly not the same for the different kernels. Altough depending on the choice of the shape parameter ε, they present a fast decay to zero (Gaussian), a medium decay (IMQ) and a slow decay (3MAT).
To measure the accuracy of the reproduction process, we computed the root-mean-square errors (RMSE) on an equally-spaced grid of evaluation points. The Figure 2 shows the results so obtained.

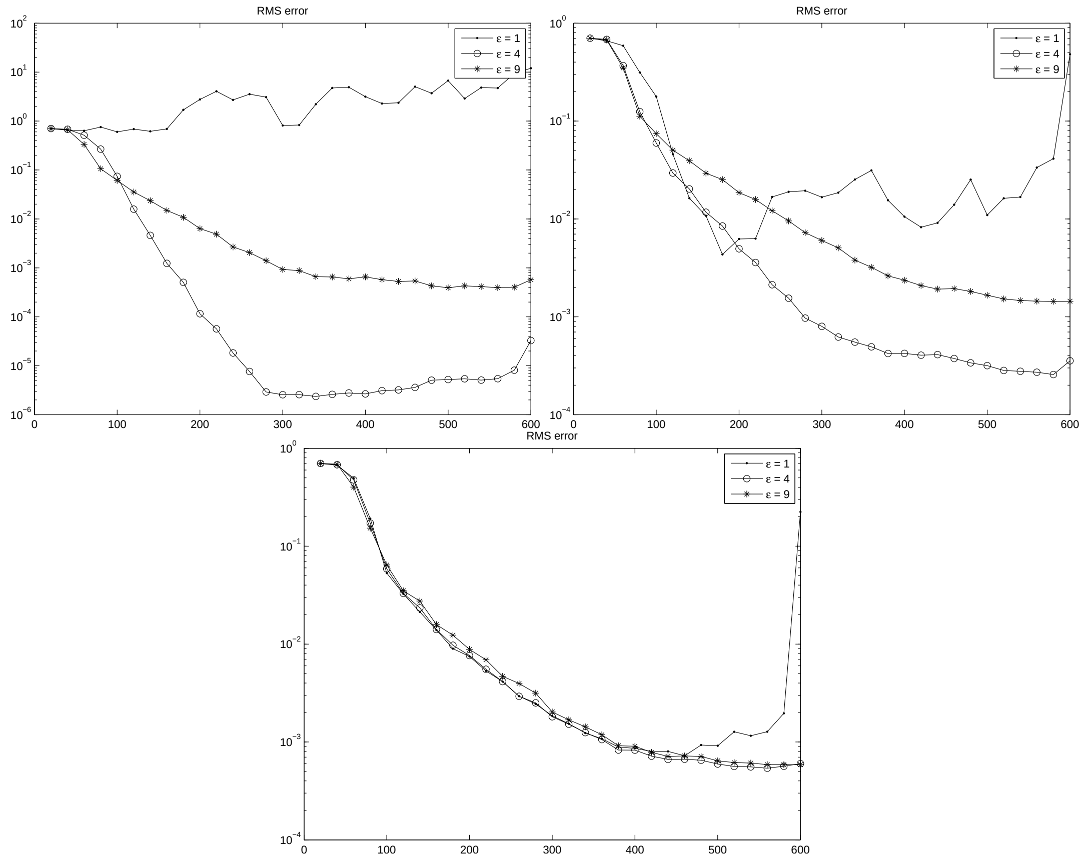

Figure 2: RMS errors for the reconstruction of f_o on Ω_1 using $\Lambda_M[f]$ for different values of M and different shape parameters, using the gaussian kernel (top left), the IMQ (top right) and the 3MAT kernel (bottom).

These results reflect the observation made on the eigenvalues. Indeed, we can see that for the 3MAT kernel the interpolant remains stable for each ε (except for the last iterations with $\varepsilon = 1$), the IMQ becomes unstable for $\varepsilon = 1$, while the gaussian presents some instability also with $\varepsilon = 4$. When a weighted least-squares approximant $\Lambda_M[f]$ for some M is used in the unstable cases we see a clear gain using a truncated version of the interpolant. Table 5.1 shows the index M such that $\Lambda_M[f]$ provides the best approximation of f_o with weighted least-squares approximation.

	$\varepsilon = 1$	$\varepsilon = 4$	$\varepsilon = 9$
Gaussian	100	340	500
IMQ	180	580	580
3MAT	460	560	580

Table 1: Optimal M for the three different kernels and shape parameters $\epsilon = 1, 4, 9$. These M correspond to the indexes such that the weighted least-squares approximant $\Lambda_M[f]$ provides the best approximation to the test function f_o on Ω_1.

As expected from the distribution of the eigenvalues of the gaussian kernel with $\varepsilon = 1$, the reducing process is not enough to avoid instability. In fact, for this choice of ε, the eigenvalues are almost all under the machine precision. Moreover, for $\varepsilon = 1$ the gaussian becomes too flat, and then there is no hope to reconstruct an oscillatory function. This is the typical behavior well-known in the flat-limit case (cf. [5]).

5.2. Comparison with the standard basis

As a second experiment we compare the approximations obtained by using the standard basis of translates with respect to the weighted least-squares one. In this example we try to reconstruct the Franke function f_F with the IMQ-kernel on the lens Ω_2 defined as the intersection of two disks with centers $C = \left(-\sqrt{2}/2, 0\right)$ and $c = \left(\sqrt{2}/2, 0\right)$ and radii $R = r = 1$ (see Figure 1).
The test compares the results obtained with the interpolant based on the standard basis and the new basis, centered on an equally-spaced grid and on a trigonometric-gaussian set of points, respectively. The reconstruction is then repeated for different shape parameters $\varepsilon = 1, 4, 9$ and for data sets $X_N \subset \Omega_2$ with $N = |X_N| < 1000$. The corresponding RMSEs are plotted in Figure 3.

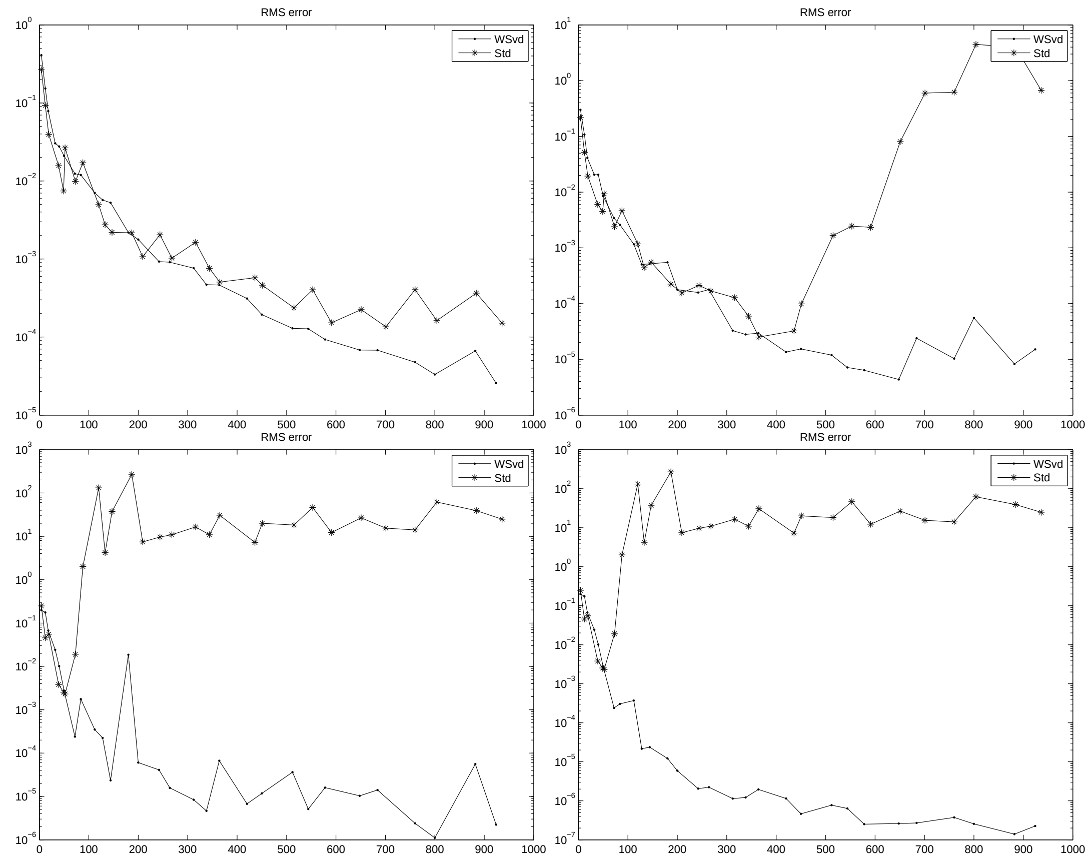

Figure 3: RMS errors for the reconstruction of f_F on the lens Ω_2 using the IMQ kernel with the standard basis and the new basis. The plots were made with different shape parameters. Top left: $\varepsilon = 9$. Top right: $\varepsilon = 4$. Bottom left: $\varepsilon = 1$. Bottom right: test with $\varepsilon = 1$ by using the interpolant based on the standard basis and the weighted least-squares approximant $\Lambda_M[f]$ with M such that $\sigma_M < 1.0e-17$.

We see that in the stable case, namely for $\varepsilon = 9$, there is only a small difference between the two bases, although for $N > 500$ the standard interpolation does not gain in accuracy. For the values $\varepsilon = 1, 4$, although for small data sets X_N the two bases does not behave so differently, when N becomes bigger the standard basis becomes unstable and growing the data-sites set does

not lead to a more accurate reconstruction. On the other hand, the interpolant based on the new basis presents a convergent behavior for each shape parameter, even if the rate of convergence is influenced by a proper choice of ε. This feature can be useful since, at least in the considered cases, there is no need to choose a particular ε to guarantee convergence, even if it is slow.
Furthermore, when a small shape parameter influences too much the stability of the interpolant, we can use instead the reduced weighted least-squares approximant $\Lambda_M[f]$. The approximation process for $\varepsilon = 1$ is repeated using $\Lambda_M[f]$ instead of $P_X[f]$, with M such that $\sigma_M < 1.0e - 17$. This threshold is choosen in order to have the best trade-off between convergence and stability. The result is shown in Figure 3 (bottom right). The approximant is clearly more stable with the same convergence rate.

As a final example, we repeated a similar comparison trying to improve the stability of the standard basis. The idea is to use an optimized shape parameter ε^* which guarantees the best possible stability. In practice we fixed the data set, the kernel and the test function, trying to find the parameter that minimizes the residual error. This has been done by the so-called *leave-one-out* cross validation strategy (cf. [7, §17.1.3]). The idea is to compute the interpolant $P_X[f]$ on the full set $X \subset \Omega$ and the N interpolants $P[f]_i$ on the reduced sets $X_i = X \setminus \{x_i\}\ \forall\ i \in \{1, \dots, N\}$, for different shape parameters $\varepsilon \in E$, $E \subset \mathbb{R}$, and then to choose the optimal ε^* defined as

$$\varepsilon^* = \arg\min_{\varepsilon \in E} \max_{i=1,\dots,N} |P_X[f](x_i) - P[f]_i(x_i)|$$

We remark that this optimization is quite expensive in terms of computational time, and can not be performed once for all, but has to be repeated if the data-sites increase. To check this optimization approach, on the square $\Omega_3 = [0,1]^2$ we considered as test function an element of the native space of the gaussian kernel $\Phi_4(x,y) := \exp(-4^2\|x-y\|_2^2)$, namely the function

$$f_{\mathcal{N}}(x) = -2\Phi_4(x,(0.5,0.5)) + \Phi_4(x,(0,0)) + 3\Phi_4(x,(0.7,0.7)) \quad x \in \Omega_3\,.$$

The RMS errors are plotted in Figure 4, using equally-spaced points and Halton points as centers of the standard basis of translates. As centers for the new basis in both cases we used the product Gauss-Legendre points. It is clear that a good choice of the shape parameter reduces the instability of the standard interpolant, altough it does not suffice to avoid it completely.
The stability of the new basis, together with the truncation strategy at M such that $\sigma_M < 1.0e-17$, allows to use the "right" shape parameter for each number of centers, and this leads to an approximant that converges to the sampled function with a tolerance near to the machine precision. Table 5.2 shows the RMS errors for different numbers of data sites, together with the optimal parameter ε^* selected by the leave-one-out cross validation.

N	196	324	529	729	900
Std - e	$1.05 \cdot 10^{-7}$	$5.23 \cdot 10^{-10}$	$3.17 \cdot 10^{-12}$	$7.15 \cdot 10^{-12}$	$1.30 \cdot 10^{-12}$
ε^*	3.75	3.81	3.84	3.87	3.98
Std - H	$3.30 \cdot 10^{-7}$	$9.31 \cdot 10^{-9}$	$7.12 \cdot 10^{-11}$	$1.56 \cdot 10^{-11}$	$1.59 \cdot 10^{-12}$
ε^*	3.75	3.84	3.89	3.92	3.95
W-Svd	$7.37 \cdot 10^{-8}$	$2.23 \cdot 10^{-11}$	$3.48 \cdot 10^{-15}$	$6.08 \cdot 10^{-15}$	$6.37 \cdot 10^{-15}$

Table 2: RMS errors for the approximation described in the last example, obtained using the new basis (W-Svd), the standard basis with optimal ε^* centered on equally-spaced points (Std - e) and on the Halton points (Std - H).

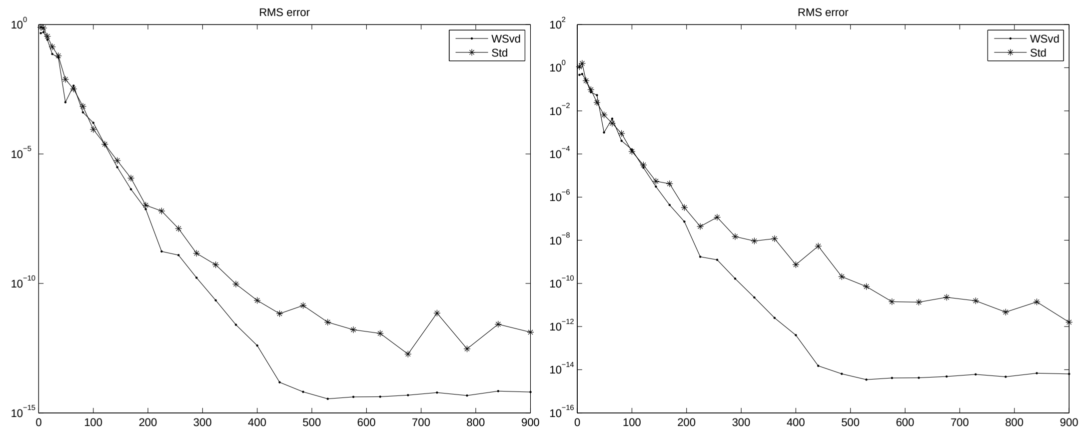

Figure 4: RMS errors for the reconstruction of $f_{\mathcal{N}}$ on the square Ω_3 using the gaussian kernel with the standard basis of translates with optimal shape parameter $\varepsilon*$ and the new basis with $\varepsilon = 4$. The standard interpolant is computed using equally spaced points (on the left) and Halton points (on the right). Our new basis is truncated at M such that $\sigma_M < 1.0e - 17$

6. Conclusions and further work

We have presented a way to construct a new kind of stable basis for radial basis function approximation. Inspired by the results presented in [12], the new basis shows properties related to its Φ-orthonormality. The particular approach used relates the discrete basis with the continuous "natural" one described in the Theorem 1, allowing to connect some functional properties of the kernel to the approximant itself.

In this setting, a more deep study could lead to a stronger use of the information provided by the kernel and its domain of definition. In particular the convergence estimate of Proposition 17 can be refined considering the rate of convergence to zero of the eigenvalues of the operator T_Φ and the property and the convergence rate of the Nyström method based on the definition of the problem, namely the cubature rule, the kernel Φ, the shape parameter ε and the set Ω.

Concerning the stability, the experiments presented in Section 5 confirm the good behavior expected from the results of Proposition 16. In particular, the new basis allows to treat approximations based on a relatively big number of points also for not optimized shape parameters and on quite general sets. This feature can be enforced thanks to the possibility to compute a weighted least-squares approximant simply truncating the interpolant. From a numerical point of view this procedure can be accomplished without thinning the data sites $X \subset \Omega$, but simply checking if the singular values of the weighted kernel matrix decay under a certain tolerance. This corresponds to solve the linear system related to the kernel matrix with a (weighted) total least-squares algorithm.

The dependence of the basis on the singular value decomposition does not allow to produce an adaptive algorithm, but forces the computation of a full factorization of the matrix for each fixed point distribution. In this sense, it would be interesting to adapt our method to work with approximations based on compactly supported kernels. Indeed, although it is possible to use them as any other kernel, a more specific implementation could benefit from the compact support structure (i.e. dealing with sparse matrices). In this setting there are eigenvalue algorithms optimized for finding only a small subset of the full spectrum of a matrix, so that it would be possible to compute an approximant based only on eigenvalues up to a certain tolerance.

Acknowledgments. We are grateful to Prof. Robert Schaback for his kindness in supporting our research and fruitful discussions. We also thank the unknown referees for their suggestions that made the paper more clear and readable.

References

[1] K. Atkinson and W. Han, *Theoretical numerical analysis: a functional analysis framework*, Springer Texts in Applied Mathematics, vol. 39, 2001.

[2] Beatson, R. K. and Levesley, J. and Mouat, C. T., *Better bases for radial basis function interpolation problems*, J. Comput. Appl. Math. 236(4) (2011), 434–446.

[3] L. Bos and M. Vianello, *Subperiodic trigonometric interpolation and quadrature*, Appl. Math. Comput. 218 (2012), 10630–10638.

[4] Stefano De Marchi and Robert Schaback, *Stability of kernel-based interpolation*, Adv. Comput. Math. 32(2) (2010), 155–161.

[5] T. A. Driscoll and B. Fornberg, *Interpolation in the limit of increasingly flat radial basis functions*, Comput. Math. Appl. 43(3-5) (2002), 413–422.

[6] B. Fornberg, E. Larsson, N. Flyer *Stable computations with Gaussian radial basis functions*, SIAM J. Sci. Comput. 33 (2) (2011) 869–892

[7] G. Fasshauer, *Meshfree approximation methods with matlab*, World Scientific, Interdisciplinary Mathematical Sciences - Vol. 6, 2007.

[8] G. Fasshauer, M. J. McCourt *Stable evaluation of Gaussian radial basis functions interpolants*, SIAM J. Sci. Comput. 34 (2) (2012) A737–A762.

[9] G. Fasshauer and J. Zhang *Preconditioning of Radial Basis Function Interpolation Systems via Accelerated Iterated Approximate Moving Least Squares Approximation*, in Progress on Meshless Methods, A. J. M. Ferreira et al. (eds.), Springer Science (2009), 57–75.

[10] Mercer J., *Functions of positive and negative type and their connection with the theory of integral equations*, Phil.Trans. Royal Society A 209 (441458) (1909), 415–446.

[11] S. Müller and R. Schaback, *A Newton basis for kernel spaces*, J. Approx. Theory 161 (2009), 645–655.

[12] M. Pazouki and R. Schaback, *Bases for Kernel-Based Spaces*, J. Comput. Appl. Math. 236 (2011), 575–588.

[13] G. Santin, *A new stable basis for RBF approximation*, Master's thesis, available online at `http://arxiv.org/abs/1210.1682`.

[14] M. Vianello and G. Da Fies, *Algebraic cubature on planar lenses and bubbles*, Dolomites Res. Notes Approx. DRNA 5 (2012), 7–12.

[15] H. Wendland, *Scattered data approximation*, Cambridge University Press, 2005.

[16] Software page of the Padova-Verona research group on "Constructive Approximation and Applications", `http://www.math.unipd.it/~marcov/CAAsoft.html`.

[17] Reference page for the *Mex* interface, `http://www.mathworks.com/help/matlab/creating-c-c-and-fortran-programs-to-be-callable-from-matlab-mex-files.html`.

[18] Home page of the *Eigen* project `http://eigen.tuxfamily.org`.